SUPPLEMENTARY MATERIAL:

Inference for Discretely Observed Stochastic Kinetic Networks with Applications to Epidemic Modeling

BOSEUNG CHOI and GRZEGORZ A. REMPALA*

Department of Computer Science and Statistics

Daegu University, Gyeongbuk, 712-714, Rep. of Korea

and

Department of Biostatistics and the Cancer Center,

Georgia Health Sciences University, Augusta, GA, 30912, USA

cbskust@gmail.com, grempala@georgiahealth.edu

A. STOCHASTIC KINETIC NETWORKS

The simplest stochastic kinetic network (SKN, or a *reaction network* in biochemical literature) is a stochastic dynamical system involving multiple species and their interactions. Here we consider only the simplest case (McQuarrie, 1967; Zacharof and Butler, 2004) which treats the system as a continuous time Markov chain whose state X is an m-vector giving the number of individuals (units) of each species present, with each reaction/interaction modeled as a transition for the state. The kth interaction is then de-

*To whom correspondence should be addressed.

termined by a vector of inputs ν_k and outputs ν'_k, specifying, respectively, the number of units of each species consumed and created, and a function of the state $h_k(x)$ that gives the rate at which the interaction occurs. Specifically, if the interaction occurs at time t, the new state becomes $X(t) = X(t-) + \nu'_k - \nu_k$, and the number of times that the kth interaction occurs by time t is given by the counting process satisfying $R_k(t) = Y_k(\int_0^t h_k(X(s))ds)$ where the Y_k are independent unit Poisson processes. Note that writing R_k in this form makes it clear why h_k is referred to as a rate, since it is indeed a rate of the corresponding Poisson process. The state of the system then satisfies the following trajectory equation (see, e.g., Andersson and Britton, 2000, Chapter 5)

$$\begin{aligned} X(t) &= X(0) + \sum_k R_k(t)(\nu'_k - \nu_k) = X(0) + \sum_k Y_k(\int_0^t h_k(X(s))ds)(\nu'_k - \nu_k) \\ &= X(0) + (\nu' - \nu)R(t), \end{aligned} \tag{A.1}$$

where ν' is the matrix with columns given by the ν'_k, ν is the matrix with columns given by the ν_k, and $R(t)$ is the vector with components $R_k(t)$.

The general law of mass action (see, e.g., Turner *and others*, 2004) states that the rate h_k is proportional to the number of distinct subsets of the species (molecules) present that can form the inputs for the reaction. Intuitively, this assumption reflects the idea that the physical system is *well-stirred* (see, e.g., Gillespie 1992; Kurtz 1981 for more precise discussions of this intuition). Defining $|\nu_k| = \sum_i \nu_{ik}$, the stochastic form of the law of mass action (Gillespie, 1992) says therefore that the Poisson rate in such system should be given by

$$h_k^M(x) = \theta_k \frac{\prod_i \nu_{ik}!}{M^{|\nu_k|-1}} \binom{x}{\nu_{1k} \cdots \nu_{mk}} = M\,\theta_k \frac{\prod_i \nu_{ik}! \binom{x_i}{\nu_{ik}}}{M^{|\nu_k|}}, \tag{A.2}$$

where M is a scaling parameter usually taken to be the total volume of the system (i.e., all species count). The values θ_k are reaction-specific kinetic constants referred to as the *network parameters*. In molecular biology or population processes modeling, these are typically the quantities which one would like to estimate using experimental data. In the sequel, and in the main body of the paper, the full vector of network parameters is denoted by $\boldsymbol{\theta}$.

Note that if M is the total (or, in some cases, maximal) size of the system by time $T < \infty$ and x gives the number of individuals of each species present, then $c = M^{-1}x$ gives the concentrations in units per unit volume. With this scaling and a large volume we have

$$h_k^M(x) \approx M\theta_k \prod_i c_i^{\nu_{ik}} \equiv M\tilde{h}_k(c). \tag{A.3}$$

Since the law of large numbers for the Poisson process implies $M^{-1}Y(Mu) \approx u$, (A.3) implies $C(t) = M^{-1}X(t) \approx C(0) + \sum_k \int_0^t \theta_k \prod_i C(s)_i^{\nu_{ik}}(\nu'_k - \nu_k)ds$, which in the large volume limit ($M \to \infty$) gives the classical law of mass action ODE

$$\dot{C}(t) = \sum_k \theta_k \prod_i C(t)_i^{\nu_{ik}}(\nu'_k - \nu_k). \tag{A.4}$$

Recent attempts to model chemical reactions taking place in biological cells as well in analyzing early stages of epidemics Ball *and others* (2006); O'Neill and Roberts (1999), have led to renewed interest in stochastic models following (A.1) (see, e.g., Wilkinson 2009). In such models the chemical notation for interactions in discrete population processes is often used where e.g., $A + B \xrightarrow{h} C$ is interpreted as *"a unit of species A combines with a unit of species B to create a unit of species C at rate $h = h(A, B, \theta)$"*. In what follows we consider a *network* of interactions (reactions) involving m popula-

tion species, $A_1, \ldots, A_m$.

$$\sum_{i=1}^{m} \nu_{ik} A_i \xrightarrow{h_k} \sum_{i=1}^{m} \nu'_{ik} A_i \qquad k = 1, 2, \ldots \tag{A.5}$$

where the ν_{ik} and ν'_{ik} are nonnegative integers.

Note that in the above SKN models we tacitly assume that the populations of interest are both spatially and phenotypically homogeneous i.e., they constitute a well-stirred physical system (see, e.g., Wilkinson, 2009). Whereas these assumptions are not expected to be satisfied in realistic scenarios, some recent analysis of the US spatial disease transmission patterns (via the so-called mobility network models) have indicated that e.g., at the onset of the epidemic they frequently may be acceptable (Mills *and others*, 2004; Balcan *and others*, 2009). We also note that it is often possible to reduce a non-homogenous system to a homogenous one via the *compartmentalization* method, the technique which is routinely used for modeling stochastic spatial features of complex biological systems (Gibson and Renshaw, 1998), also in the context of spatial epidemic models (see, e.g., Hohle *and others* (2005), Streftaris and Gibson (2004) for within-herd dynamics of disease, or recently Kouyos *and others* (2010) for changes in spatial HIV transmissions in Switzerland).

B. INFERENCE FOR COMPLETE DATA

For completeness, we present here the basic justification for the Bayesian inference procedure in case when the entire trajectory is observed. This complete data inference step is needed to construct the Gibbs sampler algorithm (Algorithm 1) in the paper.

Assume that the entire trajectory $\boldsymbol{X} = \{X(t)\}_{t=0}^{T}$ of the SKN is observable, that

is, we know both the timings and types of *all* the interactions between the species in the network up to a fixed time point $T < \infty$. Additionally, assume that the SKN has u species and v reactions with hazards $h_1(x, \theta_1), \ldots, h_v(x, \theta_v)$ and associated parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_v)$. Let r_{si} be the number of reactions of type $s,\ s = 1, 2, \ldots, v$ in the time interval $(i, i+1]$, set $n_i = \sum_{s=1} r_{si}$ as well as denote by t_{ij} and k_{ij}, respectively, the time and type of the reaction in $(i, i+1]$, $j \geqslant 1$. To construct the likelihood function, we consider the (conditional) transition density function for jth event, which, under our Markov model (in view of the well known result about arrivals from two independent Poisson variables), consists of an independent pair of event's time and type (this representation leads to the popular "direct" Gillespie algorithm, Gibson and Bruck 2000). To simplify notation, assume T to be an integer. Then, the complete likelihood function for $\boldsymbol{\theta}$ may be written as (cf., e.g., Boys *and others*, 2008)

$$L(\boldsymbol{\theta}|\boldsymbol{X}) = \prod_{i=1}^{T} \prod_{j=1}^{n_i} h_{k_{ij}}(x(t_{ij}), \theta_{k_{ij}}) \exp\left(\int_0^T h_0(x(t), \theta) dt \right) \tag{B.1}$$

where $h_0(x(t), \theta) = \sum_{s=1}^{v} h_s(x, \theta_s)$. Since, by the law of mass action, for fixed M

$$h_s(x, \theta_s) = \theta_s g_s(x) \quad s = 1 \ldots, v \tag{B.2}$$

where $g_s(\cdot)$ is a known (rational) function free of $\boldsymbol{\theta}$, we note a useful factorization of the

likelihood (B.1)

$$
\begin{aligned}
L(\boldsymbol{\theta}|\boldsymbol{X}) &= \left(\prod_{i=1}^{T} \prod_{j=1}^{n_i} \theta_{k_{ij}} g_{k_{ij}}(x(t_{ij})) \right) \exp\left(- \int_0^T \sum_{s=1}^{v} \theta_s g_s(x(t)) dt \right) \\
&\propto \left(\prod_{s=1}^{v} \theta_s^{\sum_{i=1}^{T} r_{si}} \right) \exp\left(- \sum_{s=1}^{v} \int_0^T \theta_s g_s(x(t)) dt \right) \\
&= \prod_{s=1}^{v} \theta_s^{\sum_{i=1}^{T} r_{si}} \exp\left(- \theta_s \int_0^T g_s(x(t)) dt \right) \\
&= \prod_{s=1}^{v} L_s(\theta_s|\boldsymbol{X}), \qquad \text{(B.3)}
\end{aligned}
$$

where the last equation defines quantities $L_s(\theta_s|\boldsymbol{X})$. It follows that the likelihood inference may be carried out for each θ_s separately, by maximizing the corresponding $L_s(\theta_s|\boldsymbol{X})$ quantity which is proportionate to the gamma $\Gamma(r_s + 1, \int_0^T g_s(x(t)))$ density, where $r_s = \sum_{i=1}^{T} r_{si}$. By differentiating with respect to each θ_s, we obtain their unique MLEs at the modes of the corresponding marginal gamma densities

$$
\hat{\theta}_s = \frac{\sum_{i=1}^{T} r_{si}}{\int_0^T g_s(x(t)) dt} = \frac{r_s}{\int_0^T g_s(x(t)) dt}, \quad s = 1, \ldots, v. \qquad \text{(B.4)}
$$

Note that each $\hat{\theta}_s$ may be also interpreted as the method of moments estimator, i.e., the solution of the estimating equation $\int_0^T dN_s(t) - \theta_s \int_0^T g_s(x(t)) dt = 0, \quad s = 1, \ldots, v$ where $N_s(t)$ is the s-th reaction counting process (see, e.g., Andersson and Britton 2000 Chapter 9, for more details). Due to the form of the likelihood function (B.3), a family of independent gamma distributions arrives naturally as a set of conjugate priors, that is, we take $\theta_s \sim \Gamma(a_s, b_s)$, $s = 1, \ldots, v$. Under this family of priors the application of Bayes' theorem produces posterior gamma distributions which retain independence, i.e.,

for $s = 1, \ldots, v$ we have

$$\theta_s | \boldsymbol{x} \sim \Gamma \left(a_s + r_s, b_s + \int_0^T g_s(x(t)) dt \right), \quad s = 1, \ldots, v. \tag{B.5}$$

Note that the maximum aposteriori (MAP) estimator is simply the mode of posterior distribution i.e., an adjusted MLE. In particular, in case of the *uninformative and improper* priors with $a_s = 1$ and $b_s = 0$, MLE and MAP estimator coincide. The relation (B.5) was also noted in Boys *and others* (2008).

C. Comparative Simulation Study: SIRS Model

In order to assess the performance of our inference method, we have conducted extensive simulation studies based on the synthetic data generated from the SIRS model with $T = 30$, $Y(0) = 1$, and varying M and $\boldsymbol{\theta}$. The detailed description of the five different data collection schemes is given in the main body of the paper. In addition to summarizing the results of the inference via the uniformization-based Gibbs sampler, we have also included for comparison the summary of the results obtained via the MHA-based Gibbs sampler, as described in Boys *and others* (2008) and implemented in the *StochInf* software (see, Wilkinson 2006, Chapter 10.3).

C.1 *Comparisons with Completely Observed Species*

In the main body of the paper we have presented the results for a typical trajectory of the SIRS model with $M = 25$. In Table C.I we present a more extensive set of results, for the SIRS trajectories with $M = 50$ and $M = 100$. These results are selected as a representative sample of the results of both inference method performances across the

Table C.I. *Posterior means (standard deviations) of the SIRS model parameters with $M = 50, 100$ for the uniformization Gibbs sampler (U) and the Gibbs sampler with nested Metropolis-Hastings step (MH) for different data collection scenarios and data interval lengths (m).*

	Sampler Type	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
$M = 50$				
First 10 pts ($m = 1$)	U	0.006(0.002)	0.115(0.067)	0.001(0.012)
	MH	0.006(0.002)	0.109(0.064)	0.001(0.006)
First 20 pts ($m = 1$)	U	0.009(0.002)	0.169(0.042)	0.056(0.039)
	MH	0.008(0.002)	0.157(0.039)	0.045(0.032)
All 30 pts ($m = 1$)	U	0.008(0.001)	0.194(0.038)	0.095(0.027)
	MH	0.008(0.001)	0.181(0.033)	0.081(0.023)
Sparse 10 pts ($m = 3$)	U	0.008(0.002)	0.181(0.045)	0.092(0.038)
	MH	0.022(0.004)	0.49(0.107)	0.226(0.087)
Sparse 15 pts ($m = 2$)	U	0.008(0.002)	0.19(0.045)	0.095(0.035)
	MH	0.015(0.003)	0.352(0.078)	0.167(0.061)
$M = 100$				
First 10 pts ($m = 1$)	U	0.01(0.001)	0.197(0.032)	0.062(0.042)
	MH	0.009(0.001)	0.187(0.029)	0.044(0.034)
First 20 pts ($m = 1$)	U	0.009(0.001)	0.215(0.026)	0.074(0.015)
	MH	0.009(0.001)	0.211(0.024)	0.071(0.015)
All 30 pts ($m = 1$)	U	0.009(0.001)	0.211(0.023)	0.094(0.013)
	MH	0.008(0.001)	0.182(0.02)	0.069(0.011)
Sparse 10 pts ($m = 3$)	U	0.009(0.001)	0.196(0.023)	0.085(0.013)
	MH	0.018(0.002)	0.429(0.05)	0.136(0.025)
Sparse 15 pts ($m = 2$)	U	0.01(0.001)	0.215(0.025)	0.095(0.014)
	MH	0.013(0.002)	0.306(0.038)	0.099(0.023)
True Values		0.01	0.2	0.1

multiple simulations we have conducted. The reported values are based on the 5000 steps of the Gibbs samplers after 1000 burn-in period, with the samplers convergence assessed via the usual Gelman-Rubin statistic R_{GR} as given in Gelman and Rubin (1992) with the standard stopping criterion taken as $R_{GR} \leqslant 1.1$. Whereas this criterion seemed appropriate most of the time, in some cases the comparison required a more restrictive criterion and a larger number of iterations, as detailed in the next subsection. The non-informative, improper priors $a_s = 0.1$ and $b_s = 0.1$ were used for all θs, in which case the posterior means of the marginals coincided approximately with the MLEs. As we may see, similarly to the case $M = 25$ discussed in the paper, the overall performance of the samplers is similar to each other when the observation grid is dense ($m = 1$)

and quite satisfactory (estimates have low bias and variance) when the grid is uniformly distributed across the time interval $(0, T]$. When the observed data are too concentrated in the particular trajectory segment, the effect of a temporal bias is clearly visible. This may be seen, for instance, in the significant shrinkage of the values $\hat{\theta}_3$ for the data collected from the early trajectory (bin one in Figure 2 in Section 4 of the paper), where the conversions of removed into susceptibles are happening rarely, if at all.

When the data is collected on the sparse time grid, the uniformization method is seen to have a distinct edge over the MHA-based method based on the each sampler 6000 iterations. This seems to be primarily due to the fact that, when the hidden state space is large, the MHA-based samplers converge at a noticeably slower rate then the ones based on the uniformization. We illustrate this with an example in the next section.

C.2 *Convergence Rates*

The convergence assessment for all the chains considered in the current paper is based on the so-called Gelman and Rubin method (Gelman and Rubin, 1992; Brooks and Roberts, 1998) which analyzes multiple simulated MCMC chains by comparing the variances within each chain and the variance between chains. Large deviation between these two variances indicates nonconvergence and hence the value typically computed for monitoring convergence is Gelman and Rubin statistic R_{GR} equal to the square root of the variances ratio.

Since in all simulated examples under sparse data scenarios (see, the main paper's Table 1 and Table C.I) the large inferential discrepancy between the uniformization-based (U) and the MHA-based (HM) samplers was clearly visible, we have performed some additional analysis based on a side-by-side comparison of the respective R_{GR} statistics.

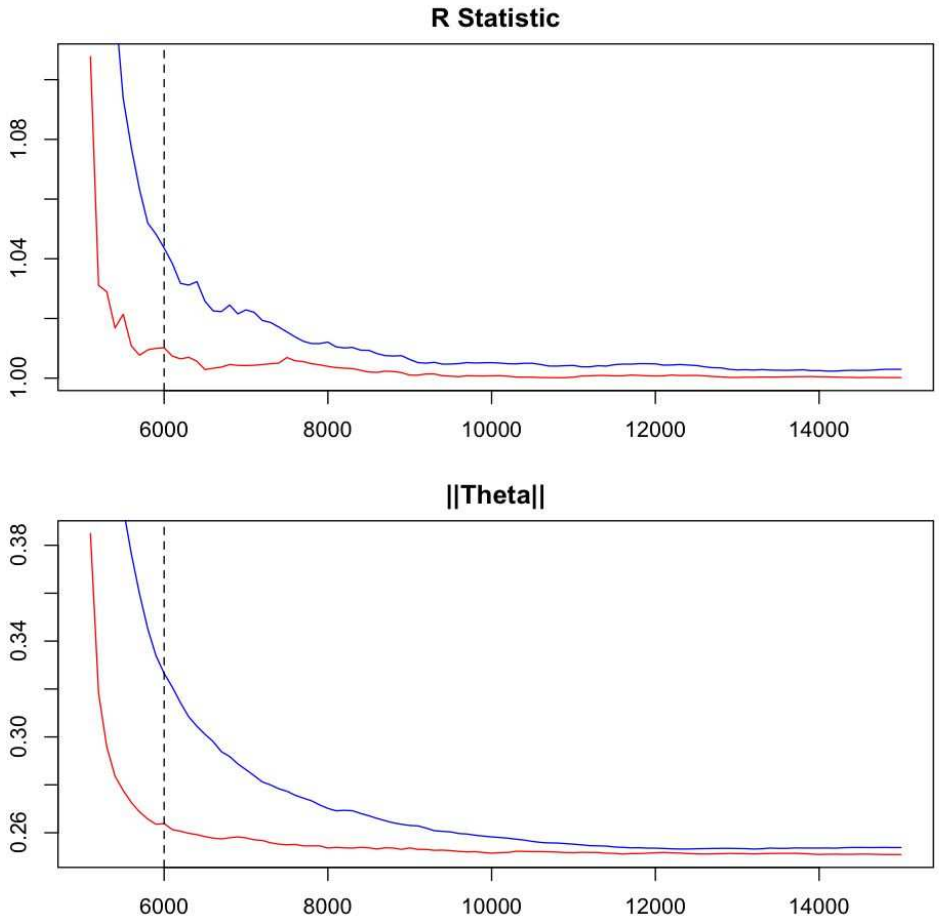

Fig. C.I. Convergence diagnostics for the Gibbs samplers results in Table 1 in the main body of the paper, under the sparse ($m = 3$, 10 data points) scenario. Top panel: The approximate value of the Rubin-Gelman R_{RG} statistics as functions of number of iterations for the MHA-based (top curve) and the uniformization (bottom curve) Gibbs samplers. Bottom panel: the values of the moving average of $||\boldsymbol{\theta}||$ (with the window size of 1000 steps) as a function of the number of iterations. The vertical line is drawn at 6000 iterations, the cut-off value used to derive the $\boldsymbol{\theta}$ estimates reported in Table 1 of the paper. The comparisons in both plots suggest the slow-convergence bias of the MHA-based sampler.

The value of R_{GR} was computed based on 5 chains initialized from various parts of the target distribution. If all chains have reached the target distribution, the posterior variance estimate should be very close to the within-chain variance and, consequently,

$R_{GR} \approx 1$.

The results are summarizrd in Figure C.I with the data from the trajectory in Figure 2 in the main body of the article and the sparse data scenario (circled dataset with $m = 3$). In the upper panel of the figure, the trajectory of the R_{RG} statistic in each sampler is plotted against the number of sampler's iterations. The corresponding values of the moving average of $||\boldsymbol{\theta}||$ from two single-trajectory samplers are plotted in the lower panel. As may be seen from the plots the MHA-based sampler needs a much larger number of iterations to converge and hence at $N = 6000$ iterations (the value used in Table 1) will typically give biased results. In contrast, the uniformization sampler converges faster and gives reliable results for $N = 6000$. Both samplers seem to give results which are very close to each other only after $N > 12000$ iterations.

C.3 *Imputation for Partially Missing Species*

Since the Algorithm 2 in Section 3.2 of the paper only estimates the number of unobserved species, the resulting uniformization Gibbs sampler converges to an approximate posterior distribution and, consequently, the inference based on combining Algorithms 1, 2 from Sections 2.1 and 3.5 of the paper is no longer exact. The effect of the approximation on the quality of the posterior estimates in the uniformization Gibbs sampler (U) is, of course, of interest. In order to illustrate the performance of the BOP-based algorithm for approximate inference in case when only partially observed species are available we have conducted additional simulation studies of the SIRS model. In these simulation scenarios we assumed that, beyond the first data point at $t = 0$, the empirical counts were available only for removed, and not for infectives or susceptibles. The results of are summarized in Table C.II below for large-to-moderate stochastic noise

scenarios $M = 25, 50$. As may be seen from the table, the approximate sampler (UB) is seen to accrue additional bias and variance due to missing information, at least in some parts of the posterior distribution. However, the effect seems to be alleviated with more timepoints added (30 in our simulation), indicating the consistency of the BOP approximation method. The convergence criteria used in reporting the simulation results are the same as those applied earlier ($N = 6000$), with the rate of convergence comparable to that illustrated in Figure C.I.

Table C.II. *Posterior means (standard deviations) of the SIRS model parameters with $M = 25, 50$ for the uniformization Gibbs sampler with BOP imputation of Algorithm 2 in Section 3.2 (UB) for different data collection scenarios and between-data interval lengths (m). Lower bias and variability is achieved with larger number of observed timepoints (results in bold), indicating the approximation consistency.*

	Sampler Type	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
$M = 25$				
First 10 pts ($m = 1$)	UB	0.015(0.009)	0.338(0.192)	0.002(0.025)
First 20 pts ($m = 1$)	UB	0.017(0.006)	0.342(0.114)	0.019(0.019)
All 30 pts ($m = 1$)	UB	**0.018(0.001)**	**0.194(0.038)**	**0.095(0.027)**
Sparse 10 pts ($m = 3$)	UB	0.012(0.003)	0.169(0.052)	0.035(0.016)
Sparse 15 pts ($m = 2$)	UB	0.015(0.004)	0.215(0.067)	0.032(0.017)
True Values		0.02	0.2	0.1
$M = 50$				
First 10 pts ($m = 1$)	UB	0.007(0.004)	0.335(0.201)	0.001(0.012)
First 20 pts ($m = 1$)	UB	0.011(0.004)	0.219(0.131)	0.013(0.018)
All 30 pts ($m = 1$)	UB	**0.011(0.004)**	**0.219(0.134)**	**0.13(0.018)**
Sparse 10 pts ($m = 3$)	UB	0.006(0.001)	0.182(0.046)	0.019(0.01)
Sparse 15 pts ($m = 2$)	UB	0.008(0.002)	0.254(0.059)	0.014(0.012)
True Values		0.01	0.2	0.1

C.4 *Marginal Plots and Diagnostics*

The marginal posterior distributions of the θ variables of interest may be obtained by considering the marginal empirical samples from the converged samplers. In our simulation experiments, in order to alleviate the effect of autocorrelation, the samples were thinned in the 5:1 proportion (after discarding the burn-in), leaving the marginal samples

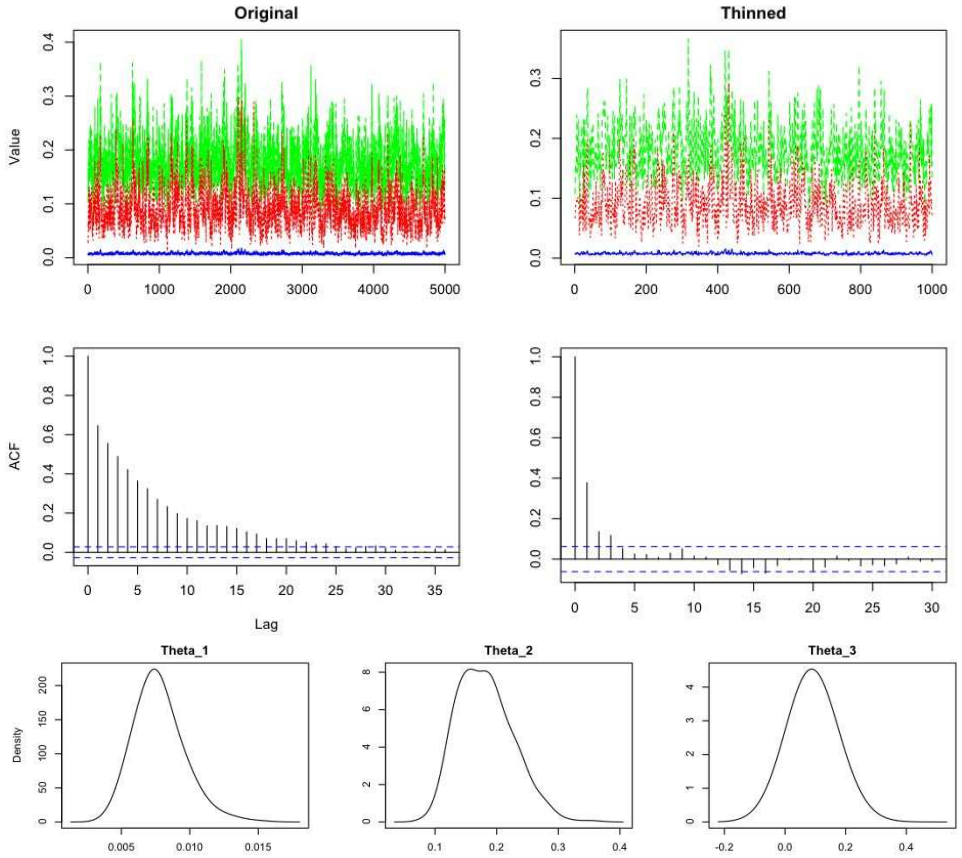

Fig. C.II. Top two rows: the trace and autocorrelation plots for the uniformization-based Gibbs sampler in the SIRS model for the sparse data (10 points, $m = 3$, $M = 50$) before and after the 5:1 thinning process. Bottom row: the marginal posterior density plots for $\theta_1, \theta_2, \theta_3$.

of size 1000 to be used for the density estimation. For illustration purposes, we present some of the diagnostic plots of the trace and autocorrelation functions before and after the thinning procedure (Figure C.II, top panels) as well as the marginal posterior density plots of the SIRS parameters $\theta_1, \theta_2, \theta_3$ (Figure C.II, bottom panels).

C.5 *Computational Considerations*

For $M = 100$ the computations for Algorithm 1 for the SIRS model took several hours to complete on the desktop computer running OS X with 3.2 Quad Core Intel Xeon processor. In general, in our simulations with SIR, SIRS and other models, we have seen that, as expected, with the increasing state space of the MCMC models the CPU requirements increased exponentially, both with the species number and the population size. For partially observed species, the use of the approximate imputation algorithm (Algorithm 2 in the main body of the paper) allowed us to roughly retain the same order of magnitude in the number the sampler steps needed for convergence, despite sampling in the extended hidden space of unobserved species. Despite all the reductions brought by the approximation, the computational cost for all UB, U and MH samplers was seen as relatively high, although in general the convergence of U and UB samplers with sparse data was achieved much faster then with MH samplers (see, Figure C.I). Albeit this is outside the scope of our current discussions, one possible way of decreasing the computational overhead for U and UB samplers, would be to appropriately adjust the value of μ given in the formula (3.1) in Section 3 of the paper, so as to efficiently reduce the size of the hidden state space which needs to be sampled by the MCMC algorithm. The appropriate adaptive algorithm for μ selection could be developed, for instance, by establishing the required exponential bounds on a jump process trajectory (given observed values) and identifying the envelopes of "high probability" trajectories along the lines of the "large deviation" principles for Markov processes (see, e.g., Kurtz and Feng 2006).

REFERENCES

ANDERSSON, H. AND BRITTON, T. (2000). *Stochastic epidemic models and their statistical analysis*. Springer Verlag.

BALCAN, D., HU, H., GONCALVES, B., BAJARDI, P., POLETTO, C., RAMASCO, J.J., PAOLOTTI, D., PERRA, N., TIZZONI, M., BROECK, W.V. *and others*. (2009). Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Medicine* **7**(1), 45.

BALL, K., KURTZ, T., POPOVIC, L. AND REMPALA, G.A. (2006). Asymptotic analysis of multiscale approximations to reaction networks. *Annals of Applied Probability* **16**(4), 1925–1961.

BOYS, R.J., WILKINSON, D.J. AND KIRKWOOD, T.B.L. (2008). Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing* **18**(2), 125–135.

BROOKS, S. AND ROBERTS, G. (1998) Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing* **8**(4), 319–335.

GELMAN, A. AND RUBIN, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–511.

GIBSON, G.J. AND RENSHAW, E. (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *Mathematical Medicine and Biology* **15**(1), 19.

GIBSON, M.A. AND BRUCK, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry Series A* **104**(9), 1876–1889.

GILLESPIE, DANIEL T. (1992). A rigorous derivation of the chemical master equation. *Physica A* **188**, 404–425.

HOHLE, M, JORGENSEN, E AND O'NEILL, PD. (2005). Inference in disease transmission experiments by using stochastic epidemic models. *Journal of the Royal Statistical Society Series C-Applied Statistics* **54**, 349–366.

KOUYOS, ROGER D, VON WYL, VIKTOR, YERLY, SABINE AND BÖNI, JÜRG. (2010, May). Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *Journal of Infectious Diseases* **201**(10), 1488–97.

KURTZ, T.G. (1981). Approximation of discontinuous processes by continuous processes. In: Arnold, L. and Lefever, R. (editors), *Proceedings, Bielefeld Conf on Stochastic Nonlinear Systems in Physics, Chemistry and Biology*. Springer-Verlag, Berlin. pp. 22–35.

KURTZ, T.G. AND FENG, J. (2006). *Large Deviations for Stochastic Processes*, Volume 131, Mathematical Surveys and Monographs. Rhode Island: AMS.

MCQUARRIE, D.A. (1967). Stochastic approach to chemical kinetics. *Journal of Applied Probability* **4**, 413–478.

MILLS, C.E., ROBINS, J.M. AND LIPSITCH, M. (2004, December). Transmissibility of 1918 pandemic influenza. *Nature* **432**(7019), 904–906.

O'NEILL, P.D. AND ROBERTS, G.O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society Series A-Statistics In Society* **162**, 121–129.

STREFTARIS, G. AND GIBSON, G.J. (2004, April). Bayesian inference for stochastic epidemics in closed populations. *Statistical Modelling* **4**(1), 63–75.

TURNER, T.E., SCHNELL, S. AND BURRAGE, K. (2004). Stochastic approaches for modelling in vivo reactions. *Computational Biology and Chemistry* **28**(3), 165–178.

WILKINSON, D.J. (2006). *Stochastic modelling for systems biology*. Chapman & Hall/CRC.

WILKINSON, D.J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics* **10**(2), 122–133.

ZACHAROF, A.I. AND BUTLER, A.P. (2004). Stochastic modelling of landfill leachate and biogas production incorporating waste heterogeneity. Model formulation and uncertainty analysis. *Waste Managment* **24**(5), 453–62.